Eurographics Workshop on 3D Object Retrieval (2014)
B. Bustos, H. Tabia, J.-P. Vandeborre, and R. Veltkamp (Editors)

Fisher encoding of adaptive fast persistent feature histograms for partial retrieval of 3D pottery objects

M. A. Savelonas[1,2] and I. Pratikakis[1,2] and K. Sfikas[2]

[1]ATHENA Research and Innovation Center, Xanthi, Greece
[2]Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi, Greece

Abstract

Cultural heritage is a natural application domain for partial 3D object retrieval, since it usually involves objects that have only been partially preserved. This work introduces a method for the retrieval of 3D pottery objects, based on partial point cloud queries. The proposed method extracts fast persistent feature histograms calculated adaptively to the mean point distances of the point cloud query. The extracted set of vectors is refined by a denoising component, which employs statistical filtering. The remaining vectors are further refined by a filtering component, which discards points surrounded by surfaces of extremely fine-grained irregularity, often associated with artefact damages. A bag of visual words scheme is used, which starts from the final set of persistent feature histogram vectors and estimates Gaussian mixture models by means of an expectation maximization algorithm. The resulting Gaussian mixture models define the visual codebook, which is used within the context of Fisher encoding. Experiments are performed on a challenging dataset of pottery objects, obtained from the publicly available Hampson collection.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—

1. Introduction

3D digitisation is considered a common practice in the cultural heritage (CH) domain, contributing on the dissemination and digital preservation of cultural thesaurus, in the form of digital libraries of CH objects. The utilitization of such libraries poses the need for partial 3D object retrieval capabilities, since very often artefacts are only partially preserved. Such a partial retrieval task is associated with three main challenges: (i) object repositories or queries may well be scanned using various types of scanning equipment or different settings, resulting in varying point cloud densities; (ii) scanned queries can be rough and noisy; (iii) it is not straightforward to effectively match a partial query against a complete 3D model, since there is a gap between their representations. This representation gap complicates the extraction of a signature that will enable a matching between partial 3D queries and complete 3D objects that share common geometric attributes.

Most partial 3D object retrieval methods are based on local shape descriptors calculated over interest points, either dense, or extracted by means of a salient point detector. A review of the state-of-the-art in this area has been included in the recent work of Li et al. [LGJ14]. Local shape descriptors are suited to partial retrieval, since a partial query and its associated complete model are intuitively expected to be similar in a local fashion. As Bronstein et al. [BBGO11] note, local feature-based methods are more diverse in the shape analysis community than in computer vision, as there is nothing equivalent to a robust feature descriptor like scale invariant feature transform (SIFT) [Low04] to be universally adopted.

There has been a considerable amount of research on local shape descriptors: starting from the spin-images introduced by Johnson and Hebert [JH99], which is a classical example of local 3D feature somehow analogous to SIFT, one can point out the normal aligned radial features (NARF) [SGVB09], which focuses on the representation of range images of scenes, and the persistent feature histogram (PFH) [RBMB08], which has originally been proposed for registration. NARF and PFH were introduced as descriptors that are directly applicable on point clouds, avoiding the need for mesh generation. They calculate and compare

DOI: 10.2312/3dor.20141051

histograms over a neighborhood centered at each point of interest. In the case of PFH and its more efficient sibling, fast PFH (FPFH) [RBB09], the histograms encode patterns of point distances within a neighborhood. This type of information is the closest one to raw point coordinates provided in point cloud input data. Still, a normal estimation algorithm is employed to calculate surface normals, which are used for the formation of PFH.

Apart from being used for the calculation of local shape descriptors, interest points also constitute an intrinsic element of the bag of visual words (BoVW) paradigm. BoVW-based methods have been applied for image classification, object detection, visual localization for robots, whereas they are an emerging trend in 3D object retrieval, with several major works appearing recently [OD11], [STP13]. In their basic form, they are distinguished by the use of an orderless collection of image features. Although such an approach lacks any structure or spatial information, it has been proved a powerful representation which achieves state of the art performance in many application domains.

In this work, a partial 3D object retrieval method is proposed, which starts from a variation of fast PFH (FPFH) that is adaptive to the mean point distances of a point cloud query. The adaptivity of FPFH to mean distance addresses the fact that point cloud densities may vary from object to object, as a result of using various types of scanning equipment, scanning from varying distances or with different settings. In addition, a statistical outlier filter refines the extracted FPFH vectors, filtering out outliers associated with noise. The FPFH vectors are further refined by a filtering component, which discards points surrounded by extremely irregular surfaces, often associated with artefact damages. The final set of FPFH vectors is used in a BoVW context wherein Fisher encoding [SPMV13] is employed. The resulting Fisher vectors are used for partial retrieval of 3D pottery objects. Such an application of a partial 3D object retrieval method on pottery objects has only appeared once in the literature, in the work of Sfikas et al. [SPK*13], which forms a baseline for the evaluation of the proposed method.

The remainder of this paper is organized as follows: in Section 2 the proposed partial 3D object retrieval method is presented. Experimental results obtained by the proposed method on a dataset of 3D pottery objects are presented in Section 3. Finally, conclusions and future perspectives of this work are discussed in Section 4.

2. Method

The proposed partial 3D object retrieval method mainly comprises two distinct components: adaptive FPFH as a 3D descriptor and Fisher encoding.

2.1. Adaptive FPFH

PFH has been proposed by Rusu et al. [RBMB08] as a pose invariant local shape descriptor, which is applicable on point clouds and represents the underlying surface model properties at a point p. It relies upon geometrical relations between k nearest neighbors, defined by means of 3D point coordinates (x,y,z) along with estimated surface normals (nx,ny,nz). However, PFH can be expanded to use additional attributes, such as curvature and second order moment invariants. In its basic form, PFH is computed as follows: i) for each point p, all of its neighbors enclosed in the sphere of a given radius r are selected (k-neighborhood), 2) for every pair of points p_i and p_j ($i \neq j$) in the k-neighborhood of p and their estimated normals n_i and n_j (p_i being the point with a smaller angle between its associated normal and the line connecting the points), a Darboux uvn frame ($u = n_i$, $v = (p_i - p_j) \times u$, $n = u \times v$) is defined and the angular variations of n_i and n_j are computed as follows: $\alpha = u \cdot n_j$, $\phi = u \cdot (p_j - p_i)/||p_j - p_i||$, $\theta = arctan(w \cdot n_j, u \cdot n_j)$.

Figure 1 presents a diagram of the influence region considered in the PFH computation for a query point p_q placed in the middle of a sphere with radius r, and all its k-neighbors are fully interconnected.

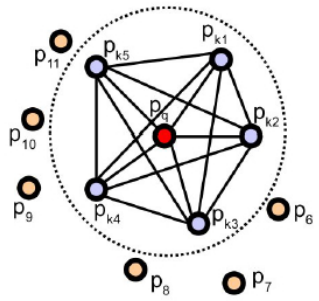

Figure 1: *The influence region diagram for a PFH. The query point (red) and its k-neighbors (blue) are fully interconnected [RBMB08].*

PFH has been reported as being several orders of magnitude slower than NARF [SGVB09]. Fast persistent feature histogram (FPFH) has been proposed by Rusu et al. [RBB09] in order to accelerate PFH computations by employing a subset of neighboring points for histogram calculation. The computational complexity of PFH for a given point cloud P with n points is O(n,k^2), where k is the number of neighbors for each point p in P. FPFH retains most of the discriminating power of PFH and has been shown to outperform spin-images in the context of registration. A further simplification introduced by Rusu et al. is to simply create d separate PFHs, one for each dimension, and concatenate them together.

A diagram of the influence region considered in the FPFH computation is illustrated in Fig. 2. For a given query point

p_q, its single persistent feature histogram (SPFH) values are first estimated by creating pairs between itself and its neighbors. This is repeated for all the points in the dataset and then the SPFH values of p_k are re-weighted using the SPFH values of its neighbors, in order to create the FPFH for p_q. As shown in the figure, some of the value pairs, which are marked with "2", will be counted twice.

The differences between PFH and FPFH are: (i) FPFH does not fully interconnect all neighbors of p_q, and is thus missing some value pairs, which might contribute to capture the geometry around p_q, (ii) PFH models a precisely determined surface around p_q, while FPFH includes additional point pairs outside the r radius sphere (though at most $2r$ away), and (iii) because of the re-weighting scheme, the FPFH combines SPFH values and recaptures some of the point neighboring value pairs. A further simplification introduced in FPFH involves the creation of separate feature histograms, one for each feature dimension, which are concatenated [RBB09].

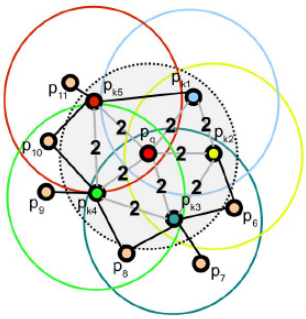

Figure 2: *The influence region diagram for FPFH. Each query point (red) is connected only to its direct k-neighbors (enclosed by the gray circle). Each direct neighbor is connected to its own neighbors and the resulting histograms are weighted together with the histogram of the query point to form the FPFH. The connections marked with "2" will contribute to the FPFH twice [RBB09].*

Addressing the issue that a dataset or a query object may well be scanned using various types of scanning equipment, from varying distances or with different settings, resulting in varying point cloud densities, the neighborhood considered in FPFH calculations is adaptively estimated for each point cloud as a linear function of the mean point distance over all k-neighborhoods, where the linear coefficient is empirically determined.

Another issue is raised by measurement errors, which may corrupt the resulting point clouds. This naturally deteriorates the quality of locally estimated point cloud features such as surface normals, leading to misleading FPFH values. The effects of such irregularities can be alleviated by performing a statistical analysis on each point neighborhood, and filtering out those points which appear as outliers. The employed outlier removal is based on the distribution of distances *between neighboring points* in the input object. For each point, the mean distance from all its neighbors is computed. All points with mean distances exceeding the standard deviation of all mean distances are considered as outliers and discarded.

Finally, points surrounded by extremely irregular surfaces, often associated with artefact damages, are naturally less distinctive with respect to the original geometry of the object. Such points tend to be surrounded by a point cloud neighborhood, for which the angles between the normal vectors take all possible values. This results in FPFH histograms with very few unoccupied bins. To cope with this issue, all FPFH vectors with less than h histogram bins of zero value, where h is empirically determined, are filtered out from subsequent calculations.

2.2. Fisher encoding

The procedure for generating a BoVW image representation can be summarized as follows: (i) a "visual vocabulary" is constructed by extracting features from all samples of a training set, followed by vector quantizing or clustering these features, where each cluster represents a "visual word" or "term". In some works, the vocabulary is called the "visual codebook". Terms in the vocabulary are the codes in the codebook, (ii) features extracted from a test sample are assigned to the closest terms in the vocabulary, using nearest neighbors or a similar clustering strategy, (iii) the occurence of each term in sample image are counted to create a normalized histogram representing a "term vector." This term vector is the BoVW representation of the sample.

In visual information retrieval, the BoVW approach defines that each sample contains a number of local visual features. Since every visual feature, or collection of similar visual features, may appear with different frequencies on each sample, matching the visual feature frequencies of two samples achieves correspondence. From the several BoVW-based methods that have been proposed for 3D object retrieval, only those of Bronstein et al. [BBGO11] and Lavoué [Lav12] have been explicitly applied on partial queries.

The BoVW component of the proposed method starts from FPFH vectors extracted for the target dataset and estimates a Gaussian mixture model (GMM) by means of an expectation maximization (EM) algorithm [MP00]. The resulting GMM defines the visual codebook, which is used within the context of Fisher encoding [SPMV13], [PD07]. This BoVW scheme provides a more general way to define a kernel from a generative processing of data and can be computed from much smaller vocabularies at a lower computational cost. It has also been recently supported in the comparative study of Chatfield et al. [CLVZ11], when compared to the basic k-means/vector quantization or the support vector encoding [ZYZH10].

Given a set of N local descriptors $\mathbf{x}_1,...,\mathbf{x}_N \in R^D$, which are use for training, a GMM $p(\mathbf{x}|\theta)$ is the probability density on R^D given by

$$p(\mathbf{x}|\theta) = \sum_{k=1}^{K} p(\mathbf{x}|\mu_k,\Sigma_k)\pi_k \quad (1)$$

$$p(\mathbf{x}|\mu_k,\Sigma_k) = \frac{1}{\sqrt{(2\pi)^D det\Sigma_k}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}-\mu_k)} \quad (2)$$

where K is the number of Gaussian components used, $\theta = (\pi_1,\mu_1,\Sigma_1,...,\pi_K,\mu_K,\Sigma_k)$ is the vector of model parameters, including the prior probability values $\pi_k \in R_+$ (which sum to one), the means $\mu_k \in R^D$, and the positive definite covariance matrices $\Sigma_k \in R^{D\times D}$ of each Gaussian component. The covariance matrices are assumed to be diagonal, so that the GMM is fully specified by $(2D+1)K$ scalar parameters. The parameters are learned by expectation maximization (EM) [MP00] from a training set of descriptors $\mathbf{x}_1,...,\mathbf{x}_N$, provided that the diagonal covariances of the components are never smaller than 0.01 times the overall diagonal covariance of the data. The GMM defines the soft data-to-cluster assignments

$$q_{ki} = \frac{p(\mathbf{x}_i|\mu_k,\Sigma_k)\pi_k}{\sum_{j=1}^{K} p(\mathbf{x}_i|\mu_j,\Sigma_j)\pi_j}, k = 1,...,K \quad (3)$$

Fisher encoding [SPMV13] captures the average first and second order differences between the local descriptors and the centres of a GMM, which can be thought of as a soft visual codebook. The construction of the encoding starts by learning a GMM model θ, as previously described. Given a set of local descriptors $\mathbf{x}_1,...,\mathbf{x}_N$ from a 3D model, let $q_{ki}, k = 1,...,K, i = 1,...,N$ be the soft assignments of the N local descriptors to the K Gaussian components, as given by (3). For each $k = 1,...,K$, define the vectors

$$\mathbf{u}_k = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik}\Sigma_k^{-1/2}(\mathbf{x}_t - \mu_k) \quad (4)$$

$$\mathbf{v}_k = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik}[(\mathbf{x}_t - \mu_k)\Sigma_k^{-1}(\mathbf{x}_t - \mu_k) - 1] \quad (5)$$

Note that, since the covariance matrices Σ_k are assumed to be diagonal, computing these quantities is quite fast. The Fisher encoding of the set of FPFH vectors is then given by the concatenation of

$$\mathbf{f}_{Fisher} = [\mathbf{u}_1^T,\mathbf{v}_1^T,...\mathbf{u}_K^T,\mathbf{v}_K^T] \quad (6)$$

Considering that both vectors, $\mathbf{u}_k$ and $\mathbf{v}_k$, have size equal to the size of the FPFH feature vector, i.e. 33, it can be derived from Eq. (6) that the resulting Fisher vector $\mathbf{f}_{Fisher}$ has size equal to $2 \times 33 \times K = 66 \times K$. Figure 3 illustrates the distinct components of the proposed pipeline for partial 3D object retrieval.

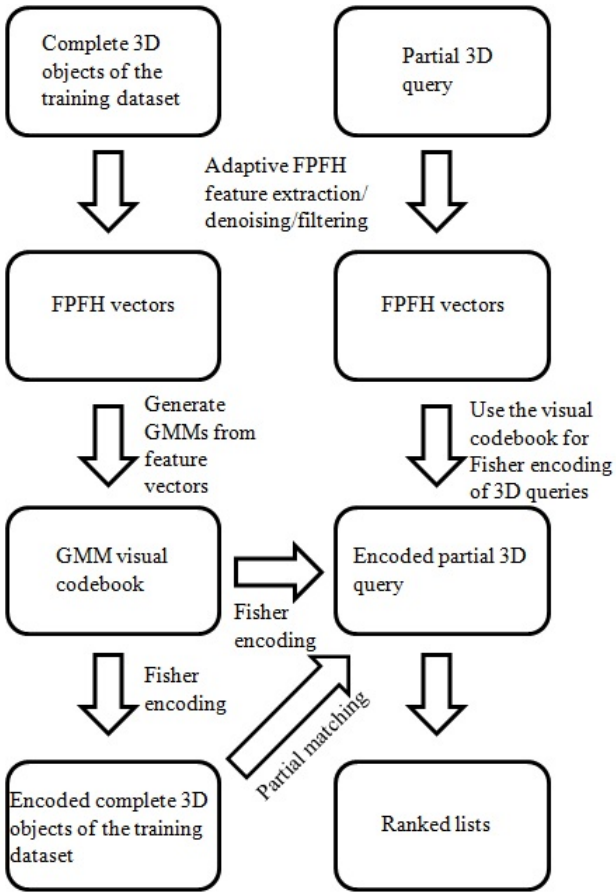

Figure 3: *The pipeline of the proposed method for partial 3D object retrieval.*

3. Experiments

Experiments have been performed to evaluate the retrieval performance of the proposed partial 3D object retrieval method and enable comparisons with the recent panoramic view-based method of Sfikas et al. [SPK*13].

3.1. Experimental setup

A 3D pottery dataset originated from the Hampson Archeological Museum collection (`http://hampson.cast.uark.edu`) is used for the experimental evaluation of the proposed method. The Hampson Archeological Museum collection composes a major source of information on the lives and history of pre-Columbian people of the Mississippi river valley.

The Centre of Advanced Spatial Technologies - University of Arkansas worked on the digitisation of numerous artefacts from the Hampson museum collection using a Konica-Minolta Vivid 9i short-range 3D laser scanner. The digitisation was performed at a precision close to 0.2 mm. The 3D digital replicas are covered by the creative common 3.0 license and are offered for online browsing or downloading in both high(>1M facets) and low (<= 25K facets) resolutions.

As a testbed for the partial 3D object retrieval experiments, 384 models of low resolution have been selected and downloaded from the website of the museum along with associated metadata information. Initially the models were classified by the museum into six general classes (bottle, bowl, jar, effigy, lithics and others). As that classification did not ensure similarities based on geometry within a given class, a futher, extended geometry-oriented classification was performed. The models where organized into thirteen classes of different populations (bottles, bowls 1 - 4, figurines, masks, pipes, tools, tripod-base vessels, conjoined vessels, twin piped vessels and others). In the sequel, five of these classes (bottle and bowls classes) were further divided into 15 subclasses, resulting in a total of 23 distinct classes. Since this dataset does not contain any partial 3D object that can be used as query, a set of 21 indicative partial queries was artificially created by slicing and cap filling an amount of complete 3D objects, originating from those classes that are densely populated. The partial queries comprise objects with a reduced surface compared to the original 3D object by a factor which ranges from 25% to 40% with a step of 5%. Although the original models of the Hampson pottery collection are represented as meshes, the proposed method is introduced to address the general case of data which are only available as point clouds and accordingly it is applied on the corresponding point clouds, without taking advantage of the original mesh normals. Instead of this, FPFH calculations involve the normals estimated by means of the least square plane fitting method of Rusu [Rus09].

The experimental evaluation employs Precision-Recall (P-R) curves and five quantitative measures: nearest neighbor (NN), first tier (FT), second tier (ST), E-measure (E) and discounted cumulative gain (DCG) [SMKF04]. For every query object that belongs to a class, recall denotes the percentage of models of the class that are retrieved and precision denotes the proportion of retrieved models that belong to the class, over the total number of retrieved models. The maximum score is 100% for both. NN is the precision at the first object of the retrieved list, when C objects have been retrieved, where C is the number of relevant objects to the query. In a similar fashion, ST is the precision when $2 \times C$ objects have been retrieved. E-measure combines precision and recall metrics into a single number and the DCG statistic weights correct results close to the top of the ranked list more than correct results later in the list, under the assumption that a user is less likely to consider the latter [SMKF04].

It should be mentioned that the Hampson collection, along with the artificially created queries and the previosuly mentioned performance measures were used in the work of Sfikas et al. [SPK*13] and they are adopted here for comparison purposes.

The linear coefficient connecting the FPFH radius and the mean point distance has beem experimentally determined as 1.7. In a similar fashion, the parameter h of the filtering component, which discards points surrounded by extremely irregular surfaces, is set to 8. The FPFH implementation uses 11 binning subdivisions for each one of α, ϕ and θ, as well as a decorrelated scheme, which results in a histogram of 33 bins. With respect to Fisher encoding, $K = 10$ GMMs were experimentally determined to be sufficient for the construction of the visual codebook, leading to Fisher vectors of $66 \times 10 = 660$ components (see subsection 2.2). Lloyds' variant [Llo82] of k-means was used as pre-clustering to initialize GMM construction. The signed square root function is applied to the resulting Fisher vectors, followed by L_2 normalization. Adaptive radius estimation, FPFH filtering components and the aforementioned configuration of Fisher encoding, have all been found to enhance the retrieval performance obtained in preliminary experiments.

The proposed method has been developed on a hybrid Matlab/C++ architecture, employing PCL [RC13] for FPFH feature calulation and VLFeat library [VF10] for Fisher encoding. The experiments have been performed on an Intel Core i5 system, operating at 2.67 GHz system with 4 GB of RAM.

3.2. Results

Table 1 presents the retrieval performance, quantified by NN, FT, ST, E-measure and DCG, which was obtained by the panoramic view-based method of Sfikas et al. [SPK*13] and the proposed 3D object retrieval method when querying the pottery dataset with partial pottery objects of various degrees of partiality (25%-40%). It can be observed that the proposed method outperforms the panoramic view-based method in most measures. In particular, it consistently performs better in terms of NN, where it achieves a nearly perfect score. This can be attributed to the fact that FPFH accurately reflects local geometry, facilitating matching of the partial query with its originating complete object, on the basis of their common parts. In addition, the proposed method performs better in terms of FT, E-measure and DCG for three out four degrees of partiality. This is reversed for ST, which is however a less indicative retrieval measure than FT.

Figure 4 illustrates the average P-R scores for the proposed method and the panoramic view-based method, obtained using queries with partiality ranging from 25% to 40%. It can be observed that the areas below the P-R curves of the proposed method are consistently larger, for all degrees of partiality. These areas can be quantified by means

of the mean average precision (MAP), which in the case of the proposed method is approximately equal to 0.38. Interestingly, retrieval performance of both methods is not a consistently increasing function of the degree of completeness of the partial query, as can be derived from the retrieval measures of Table 1 and the P-R scores of Fig. 4. Figure 5 shows example partial queries, accompanied by the top-5 retrieved template objects. It can be observed that although there is considerable variability in the pottery shapes, the proposed method succeeds in retrieving similar objects.

Table 1: *Five quantitative measures for the panoramic view-based method of Sfikas et al. [SPK*13] and the proposed partial 3D object retrieval method, obtained on the Hampson pottery collection. All measures are normalized.*

Method	NN	FT	ST	E	DCG
Pan. (25%)	0.230	0.227	0.388	0.185	0.587
Pan. (30%)	0.428	0.289	0.495	0.228	0.655
Pan. (35%)	0.619	0.372	0.536	0.327	0.713
Pan. (40%)	0.857	0.288	0.508	0.237	0.683
Prop. (25%)	0.952	0.320	0.461	0.267	0.694
Prop. (30%)	0.905	0.340	0.456	0.262	0.692
Prop. (35%)	1.000	0.331	0.450	0.260	0.689
Prop. (40%)	0.952	0.337	0.468	0.274	0.701

On a wider context, when assessing the performance of a partial 3D object retrieval method in absolute terms, one should take into account that partial retrieval remains an open and challenging research area.

The average computational time for the construction of GMMs, for the complete dataset, is approximately 1 minute, however this constitutes an offline process. The average computational time for computing the retrieval list of a single query is approximately 2 seconds.

4. Conclusions

This work introduces a partial 3D object retrieval method which is applied on 3D pottery objects and addresses the general case of data which are only available as point clouds. The proposed method employs a variant of FPFH which is adaptive to the mean point distances of a point cloud query. This adaptivity to mean distance addresses the fact that point cloud densities may vary from object to object, as a result of using various types of scanning equipment, varying scanning distances or different settings. In addition, a statistical outlier filter is employed for denoising. The remaining FPFH vectors are further refined by a filtering component, which discards points surrounded by extremely irregular surfaces, often associated with artefact damages. The final set of FPFH feature vectors is encoded by means of Fisher encoding [SPMV13]. This BoVW scheme provides a more general way to define a kernel from a generative processing of data, whereas it can be computed from much smaller vocabularies at a lower computational cost and has been supported in comparative studies over BoVW alternatives, which include the basic *k*-means/vector quantization scheme and support vector encoding.

Experimental comparisons with the panoramic view-based method of Sfikas et al. [SPK*13] on a dataset of pottery objects, obtained from the publicly available Hampson collection, lead to the conclusion that the proposed method achieves higher retrieval performance. As a final note, when assessing the retrieval performance of the proposed method, one should consider that partial 3D object retrieval is still an open and challenging research area.

Future perspectives of this work include the extension of the BoVW component to include spatial context and an additional option to use the original mesh normals in the FPFH calculations, in cases of data represented as meshes.

Acknowledgements

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 600533 PRESIOUS.

References

[BBGO11] BRONSTEIN A., BRONSTEIN M., GUIBAS L., OVSJANIKOV M.: Shape google: geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics 30*, 1 (2011), 1–20. doi:10.1145/1899404.1899405. 1, 3

[CLVZ11] CHATFIELD K., LEMPITSKY V., VEDALDI A., ZISSERMAN A.: The devil is in the details: an evaluation of recent feature encoding methods. In *Proc.BMVC'11* (2011), pp. 1–12. 3

[JH99] JOHNSON A., HEBERT M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 5 (1999), 433–449. doi:10.1109/34.765655. 1

[Lav12] LAVOUÉ G.: Combination of bag-of-words descriptors for robust partial shape retrieval. *The Visual Computer 28*, 9 (2012), 931–942. doi:10.1007/s00371-012-0724-x. 3

[LGJ14] LI B., GODIL A., JOHAN H.: Hybrid shape descriptor and meta similarity generation for non-rigid and partial 3d model retrieval. *Multimedia Tools and Applications* (2014). doi:10.1007/s11042-013-1464-2. 1

[Llo82] LLOYD S.: Least squares quantization in pcm. *IEEE Transactions on Information Theory 28*, 2 (1982), 129–136. doi:10.1109/TIT.1982.1056489. 5

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. doi:10.1023/B:VISI.0000029664.99615.94. 1

[MP00] MOOSMAN F., PEEL D.: *Finite Mixture Models*. Wiley, 2000. 3, 4

[OD11] O'HARA S., DRAPER B.: Introduction to the bag of features paradigm for image classification and retrieval. *Computing Research Repository* (2011). 2

[PD07] PERRONIN F., DANCE C.: Fisher kernels on visual vocabularies for image categorization. In *Proc.CVPR'07* (2007). 3

[RBB09] RUSU R. B., BLODOW N., BEETZ M.: Fast point feature histograms (fpfh) for 3D registration. In *Proc.ICRA'09* (2009), pp. 3212–3217. 2, 3

[RBMB08] RUSU R. B., BLODOW N., MARTON Z., BEETZ M.: Aligning point cloud views using persistent feature histograms. In *Proc.IROS'08* (2008), pp. 3384–3391. 1, 2

[RC13] RUSU R., COUSINS S.: 3D is here: Point cloud library (pcl). In *Proc.ICRA'13* (2013). 5

[Rus09] RUSU R.: *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany, 2009. 5

[SGVB09] STEDER B., GRISETTI G., VAN M., BURGARD L. W.: Robust online model-based object detection from range images. In *Proc.IROS'09* (2009), pp. 4739–4744. 1, 2

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The princeton shape benchmark. In *Proc.SMI'04* (2004), pp. 167–178. 5

[SPK*13] SFIKAS K., PRATIKAKIS I., KOUTSOUDIS A., SAVELONAS M., THEOHARIS T.: 3D object partial matching using panoramic views. In *Proc.ICIAP Workshops* (2013), pp. 169–178. 2, 4, 5, 6, 8

[SPMV13] SÁNCHEZ J., PERRONNIN F., MENSINK T., VERBEEK J.: Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision 105*, 3 (2013), 222–245. doi:10.1007/s11263-013-0636-x. 2, 3, 4, 6

[STP13] SFIKAS K., THEOHARIS T., PRATIKAKIS I.: 3D object retrieval via range image queries in a bag-of-visual-words context. *The Visual Computer 29* (2013), 1351–1361. doi:10.1007/s00371-013-0876-3. 2

[VF10] VEDALDI A., FULKERSON B.: Vlfeat - an open and portable library of computer vision algorithms. In *Proc.ACM ICM'10* (2010), pp. 1469–1472. 5

[ZYZH10] ZHOU X., YU K., ZHANG T., HUANG T.: Image classification using super-vector coding of local image descriptors. In *Proc.ECCV'10* (2010). 3

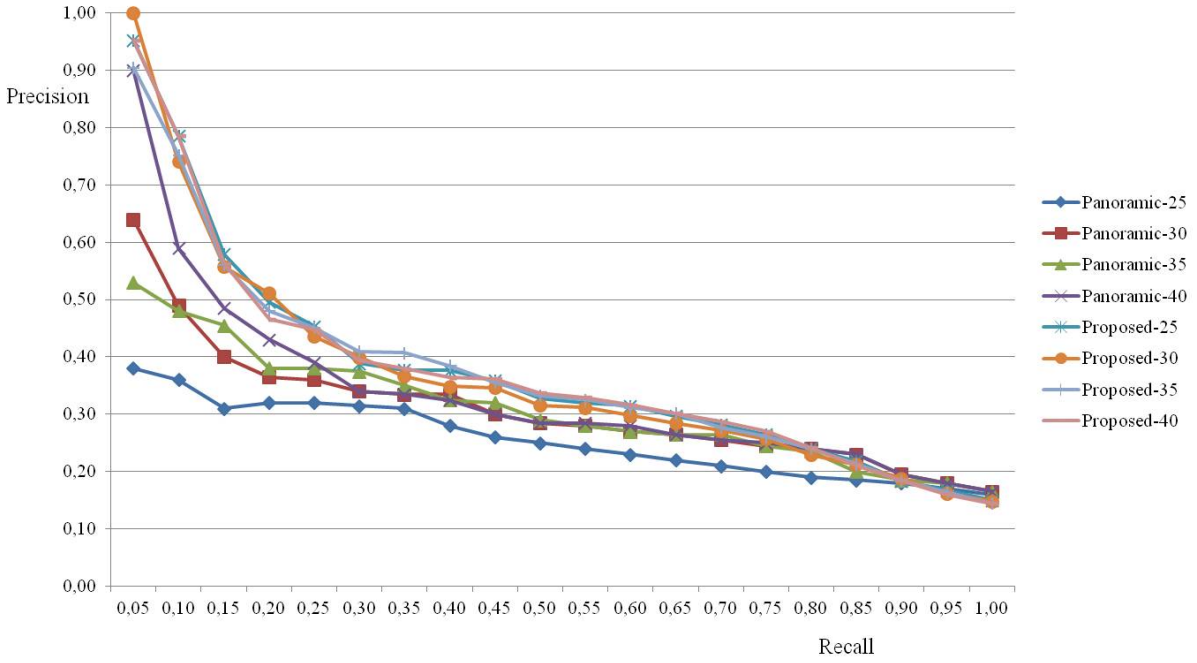

Figure 4: *Average P-R scores for the pottery dataset originating from the Hampson Archaeological Museum collection. Illustrated is the performance of the proposed method, as well as of the panoramic views-based method of Sfikas et al. [SPK*13], obtained using queries with partiality ranging from 25% to 40%.*

Figure 5: *Example retrieval results from the pottery dataset. At each row, a partial query (column 1) and a ranked list of the retrieved 3D objects (columns 2 - 6) are shown.*